XLVII Konferencja

Statystyka Matematyczna

29 listopada – 3 grudnia 2021
Będlewo

Spis treści

Część I

Wprowadzenie

Konferencja Statystyka Matematyczna odbywa się rokrocznie od wielu lat. Od 29. listopada do 3. grudnia 2021 roku odbędzie się już XLVII edycja tej konferencji, której miejscem będzie Ośrodek Badawczo-Konferencyjny Instytutu Matematycznego Polskiej Akademii Nauk w Będlewie. Konferencja ma charakter hybrydowy.

Celem konferencji jest zebranie naukowców dzielących wspólne zainteresowania związane ze statystyką matematyczną i jej zastosowaniami. Do głównych tematów poruszanych na konferencji należą:

- estymacja parametryczna i nieparametryczna,
- planowanie eksperymentów,
- liniowe i uogólnione modele liniowe,
- testowanie hipotez statystycznych,
- szeregi czasowe,
- analiza danych wielowymiarowych,
- analiza asymptotyczna,
- modele graficzne.

Planowane są również wystąpienia poświęcone metodologii analizy dużych zbiorów danych (big-data) oraz danych wysokowymiarowych (high-dimensional), ze szczególnym wykorzystaniem technik uczenia maszynowego. Konferencja ma charakter teoretyczny, jednakże rozwiązania przedstawiane w czasie wystąpień mają zastosowanie w wielu dyscyplinych współczesnej nauki i techniki, m.in. w genetyce, biologii, ekonomii, rolnictwie czy naukach inżynierskich.

Komitety i organizatory

Komitet naukowy

- prof. dr hab. Małgorzata Bogdan
- prof. dr hab. Teresa Ledwina
- prof. dr hab. Jan Mielniczuk
- prof. dr hab. Wojciech Niemiro

Komitet organizacyjny

- dr hab. inż. Katarzyna Filipiak
- dr hab. Łukasz Smaga
- dr hab. Magdalena Szymkowiak
- dr inż. Monika Mokrzycka
- mgr Adam Mieldzioc

Organizatorzy

- Instytut Matematyki, Politechnika Poznańska
- Instytut Automatyki i Robotyki, Politechnika Poznańska
- Wydział Matematyki i Informatyki, Uniwersytet im. Adama Mickiewicza, Poznań
- Katedra Metod Matematycznych i Statystycznych, Uniwersytet Przyrodniczy, Poznań
- Instytut Genetyki Roślin PAN, Poznań
- Centrum Matematyczne im. Stefana Banacha, Instytut Matematyczny PAN, Warszawa

Część II

Program konferencji

Program

Poniedziałek, 29 listopada 2021

8:00 Śniadanie

9:50–10:00 Otwarcie konferencji
10:00–11:00 *D. Klein:* Estimation and testing in multivariate linear models (I)

11:00–11:30 Przerwa

11:30–11:50 *M. Bogdan:* Pattern recovery by SLOPE when the design matrix is orthogonal
11:50–12:10 *T. Skalski:* Model recovery by SLOPE

12:10–12:30 Przerwa

12:30–12:50 *T. Rychlik:* Bounds on the expectations of L-statistics from iid symmetric samples in various scale units
12:50–13:10 *L. Pańczyk:* On the choice of the optimal linear combination of two order statistics in quantile estimation distributions with given moments
13:10–13:30 *K. Jasiński:* Conditional reliability properties of k-out-of-n system composed of different types of components

14:00 Obiad

16:00–17:00 *D. Klein:* Estimation and testing in multivariate linear models (II)

17:00–17:30 Przerwa

17:30–17:50 *M. Łazęcka:* Multiple testing of conditional independence hypotheses using information-theoretic measures
17:50–18:10 *M. Janiszewska:* Estimation of block structured covariance matrix
18:10–18:30 *M. John:* Various tests for independence under block compound symmetry covariance structure

19:00 Kolacja

Wtorek, 30 listopada 2021

8:00 Śniadanie

10:00–11:00 *D. Klein:* Estimation and testing in multivariate linear models (III)

11:00–11:30 Przerwa

11:30–11:50 *W. Niemiro:* Sensitive questions in survey statistics: some new versions of Item Count Technique

11:50–12:10 *K. Rudaś:* Shrinkage estimators for uplift regression

12:10–12:30 Przerwa

12:30–12:50 *W. Rejchel:* Selection consistency for high-dimensional categorical data

12:50–13:10 *G. Wyłupek:* A nonparametric test for paired data

13:10–13:30 *K. Bogdan:* Maximum likelihood estimation for discrete exponential families and random graphs

14:00 Obiad

16:00–17:00 *B. Miasojedow:* Computational method for high dimensional statistics (I)

17:00–17:30 Przerwa

17:30–17:50 *A. Mieldzioc:* The comparison of the estimators of linearly structured covariance matrix

17:50–18:10 *A. Geras:* Applying Metropolis-within-Gibbs algorithm to the problem of cell type deconvolution in tissues from spatial transcriptomics data

18:10–18:30 *P. Sulewski:* Two methods of conjoint summands of generating bivariate and trivariate normal pseudo-random numbers

19:00 Kolacja

Środa, 1 grudnia 2021

8:00 Śniadanie

10:00 Wycieczka

14:00 Obiad

16:00–17:00 *B. Miasojedow:* Computational method for high dimensional statistics (II)

17:00–17:30 Przerwa

17:30–17:50 *P. Pokarowski:* Objective Bayes factors in linear models
17:50–18:10 *A. Markiewicz:* Linear sufficiency and admissibility under various loss functions
18:10–18:30 *M. Kos:* Variable selection in linear model via generalized SLOPE

19:00 Kolacja konferencyjna

Czwartek, 2 grudnia 2021

8:00 Śniadanie

10:00–11:00 *B. Miasojedow:* Computational method for high dimensional statistics (III)

11:00–11:30 Przerwa

11:30–11:50 *J. Mielniczuk:* Joint estimation of posterior probability and propensity score for positive and unlabelled data
11:50–12:10 *P. Teisseyre:* Feature selection for positive-unlabelled multi-label classification
12:10–12:30 *K. Łatuszyński:* Adaptive MCMC for multimodal targets

12:30–12:50 Przerwa

12:50–13:10 *Ł. Smaga:* Permutation test for the multivariate coefficient of variation in factorial designs
13:10–13:30 *P. Sulewski:* The Weibull lifetime model with randomized failure-free time

14:00 Obiad

16:00–16:20 *A. Dudek:* Spectral density estimation for nonstationary data with nonzero mean function
16:20–16:40 *A. Szczepańska-Álvarez:* Testing of factor effects in a three level model
16:40–17:00 *M. Mokrzycka:* Discrepancy measures in the context of power study for testing covariance structures

17:00–17:30 Przerwa

17:30–17:50 *Z. Szkutnik:* Adaptivity via a discrepancy principle for Poisson inverse problems
17:50–18:10 *P. Grzegorzewski:* Goodness-of-fit tests for imprecise data
18:10–18:30 *S. Piątek:* Estimation of parameters and quantiles of the Weibull distribution

19:00 Kolacja

Piątek, 3 grudnia 2021

8:00 Śniadanie

10:00–10:20 *M. Szymkowiak:* Conditions on system signatures assuring the preservation of some distribution properties of component lifetimes by system lifetime

10:20–10:40 *A. Goroncy:* On the bounds on expected order statistics from the decreasing reversed failure rate distributions

10:40–11:00 Przerwa

11:00–11:20 *K. Furmańczyk:* Some proposal of the high dimensional PU learning classification procedure

11:20–11:40 *K. Filipiak:* Approximate normality in testing compound symmetry under large- and high-dimensional settings

11:40–11:50 Zakończenie konferencji

12:00 Obiad

Część III

Zaproszeni wykładowcy

Estimation and testing in multivariate linear models

Daniel Klein

P. J. Šafárik University, Košice, Slovakia

Abstract

Experiment is one of the main tool of scientific research and it typically results in a set of measured or observed data. There are two main problems to solve while analysing data - to determine the values of unknown parameters, and to test the hypotheses about the values of unknown parameters. The talk will be focused on the problems of estimation and testing of unknown parameters in the growth curve model and its generalized versions as well as in two-level multivariate linear models. The estimators of unknown parameters under some specific conditions, especially under assumption of special covariance matrix structure, and tests of mean and covariance structure will be studied.

Computational method for high dimensional statistics

Błażej Miasojedow

University of Warsaw

Abstract

High-dimensional statistics in recent years have posed new challenges for computational methods. On the one hand, statistical problems are the motivation for creating new algorithms, on the other hand, the analysis of numerical methods shows the directions of the development of new statistical methods. In my lecture I am going to talk about results in the field of computational statistics, focusing mainly on stochastic optimization methods.

Część IV

Streszczenia referatów

Maximum likelihood estimation for discrete exponential families and random graphs

Krzysztof Bogdan[1], **Michał Bosy**[2], **Tomasz Skalski**[1,3]

[1] Wrocław University of Science and Technology
[2] University College London, UK
[3] University of Angers, France

Abstract

We shall discuss the existence of the maximum likelihood estimator for discrete exponential families, based on preprint [1]. Our criterion is simple to apply as we show in various settings, e.g., for exponential models of random graphs. As an application, we specify the size of independent identically distributed samples for which the maximum likelihood estimator exists with high probability.

References

1. Bogdan, K., Bosy, M. and T. Skalski (2021). Maximum likelihood estimation for discrete exponential families and random graphs. *arXiv, 1911.13143* `https://arxiv.org/abs/1911.13143`

Pattern recovery by SLOPE when the design matrix is orthogonal

Tomasz Skalski[1], Małgorzata Bogdan[2], Piotr Graczyk[3], Bartosz Kołodziejek[4], Patrick Tardivel[5], Maciej Wilczyński[1]

[1] Wrocław University of Science and Technology
[2] University of Wrocław
[3] University of Angers, France
[4] Warsaw University of Technology
[5] University of Burgundy, France

Abstract

Sorted L-One Penalized Estimator (SLOPE) is a relatively new convex regularization method for fitting high-dimensional regression models. SLOPE allows to reduce the model dimension by eliminating some of the regression coefficients and by equalizing some of nonzero coefficients. This allows to identify situations, where some of true regression coefficients are equal to each other. In this talk we will introduce the SLOPE pattern, i.e. the set of relations between the true regression coefficients, which can be identified by SLOPE under certain conditions. We will also present theorem on the consistent pattern recovery by SLOPE when the design matrix is orthogonal and illustrate advantages of the SLOPE clustering in the context of the high frequency signal denoising.

Spectral density estimation for nonstationary data with nonzero mean function

Anna Dudek[1], Łukasz Lenart[2]

[1] AGH University of Science and Technology, Cracow
[2] Cracow University of Economics

Abstract

We introduce a new approach for nonparametric spectral density estimation based on the subsampling technique, which we apply to the important class of nonstationary time series. These are almost periodically correlated sequences. In contrary to existing methods our technique does not require demeaning of the data. On the simulated data examples we compare our estimator of spectral density function with the classical one. Additionally, we propose a modified estimator, which allows to reduce the leakage effect.

Keywords

Subsampling, Almost periodically correlated, Covariance matrix estimation, Mixing.

Acknowledgements

Anna Dudek acknowledges support from the King Abdullah University of Science and Technology (KAUST) Research Grant OSR-2019-CRG8-4057.2.

References

1. Dudek, A.E. and Ł. Lenart. Spectral density estimation for nonstationary data with nonzero mean function. Submitted.

Approximate normality in testing compound symmetry under large- and high-dimensional settings

Katarzyna Filipiak[1], **Daniel Klein**[2], **Jolanta Pielaszkiewicz**[3]

[1] Poznań Univeristy of Technology
[2] P. J. Šafárik University in Košice, Slovakia
[3] Linköping University, Sweden

Abstract

In this paper the Rao score and likelihood ratio tests for hypothesis related to compound symmetry covariance structure of multivariate data are studied. Under the assumption of large-dimensionality the normal approximation of the Rao score test statistics distribution is proven as well as the exact and approximate distributions of the likelihood ratio test are derived. Simulation studies show the advantage of the Rao score test over the likelihood ratio test in both studied contexts: type I error and power. Moreover, the Rao score test is available in the case of high-dimensionality, and it is shown that the normal approximation matches well its distribution in this case.

Keywords

Compound symmetry, Rao score test, Likelihood ratio test, Large-dimensionality, High-dimensionality.

Some proposal of the high dimensional PU learning classification procedure

Konrad Furmańczyk[1], **Marcin Dudziński**[1], **Diana Dziewa-Dawidczyk**[1],

Warsaw University of Life Sciences

Abstract

In [1] we propose a new classification method for positive and unlabeled (PU) data, called the LassoJoint classification procedure, which combines the thresholded Lasso approach in the first two steps with the joint method based on logistic regression, introduced by Teisseyre et. al. [2], in the last step. We prove that, under some regularity conditions, our procedure satisfies the screening property. We also conduct some simulation study in order to compare the proposed classification procedure with the oracle method. Prediction accuracy of the proposed method has been verified for some selected real datasets.

Keywords

Positive unlabeled learning, Logistic regression, Thresholded Lasso.

References

1. Furmańczyk, K., Dudziński, M. and D. Dziewa-Dawidczyk (2021). Some proposal of the high dimensional PU learning classification procedure. *Computational Sciences-ICCS 2021, LNCS*, 18–25.
2. Teisseyre, P., Mielniczuk, J. and M. Łazęcka (2020). Different strategies of fitting logistic regression for positive and unlabelled data. *Computational Sciences-ICCS 2020*, 3–17.

Applying Metropolis-within-Gibbs algorithm to the problem of cell type deconvolution in tissues from spatial transcriptomics data

Agnieszka Geras[1], Shadi Darvish Shafighi[2], Kacper Domżał[2], Igor Filipiuk[2], Łukasz Raczkowski[2], Hosein Toosi[3], Leszek Kaczmarek[4], Łukasz Koperski[5], Jens Lagergren[3], Dominika Nowis[5], Ewa Szczurek[2]

[1] Warsaw University of Technology
[2] University of Warsaw
[3] Royal Institute of Technology, Stockholm, Sweden
[4] Polish Academy of Sciences, Warsaw
[5] Medical University of Warsaw

Abstract

Markov chain Monte Carlo methods have various interesting applications to Bayesian modelling of real-world phenomena. In our work, we focus on a biological problem of spatial mapping of cell types in tissues examined using the famous spatial transcriptomics technique [2]. We apply the Metropolis-Within-Gibbs sampler to infer parameters of a novel hierarchical Bayesian model of gene expression in spatial transcriptomics data.
The spatial transcriptomics (ST) procedure (Nature's Method of the Year 2020 [1]) enables to gain an insight not only into the level of gene expression, but also enables to map this activity spatially. Importantly, ST spots (tiny areas selected on a tissue, where gene expression is measured), contain multiple cells, therefore the observed signal inevitably conveys information about mixtures of cells of different types, what poses a challenge to decompose cell-type mixtures. We present an innovative approach to solve this task, that is based on probabilistic graphical models and, unlike methods proposed so far [3,4], does not require single cell data, but instead needs additional prior knowledge on marker genes. Our novel probabilistic model called Celloscope was applied on mouse brain data and was able to successfully indicate brain structures and spatially distinguish between two main neuron types: inhibitory and excitatory. We also investigate the immune contexture of the tumour microenvironment in prostate cancer.

Keywords

Markov chain Monte Carlo, Spatial transcriptomics, Probabilistic modelling, Cell types mapping.

References

1. Marx, V. (2021). Method of the Year: spatially resolved transcriptomics. *Nat. Methods. 18*, 9–14.
2. Berglund, E., et al. (2018). Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications 9*, Article number: 2419.
3. Andersson, A., et al. (2020). Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Communications Biology 3*, Article number: 565.
4. Cable, D.M., et al. (2021). Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology* DOI: 10.1038/s41587-021-00830-w.

On the bounds on expected order statistics from the decreasing reversed failure rate distributions

Agnieszka Goroncy[1], **Tomasz Rychlik**[2]

[1] Nicolaus Copernicus University in Toruń
[2] Polish Academy of Sciences, Warsaw

Abstract

We introduce a class of distributions called the generalized negative Pareto family of distributions (GNPD). For a fixed $\alpha \in R$ and an absolutely continuous distribution function F we consider the α-generalized reversed failure rate and the respective classes of distributions with the monotone generalized reversed failure rate, which are defined in terms of the convex transform order of the considered cdf F and the GNPD distribution function. We focus on order statistics $X_{1:n}, \ldots, X_{n:n}$, based on the distribution function F with finite expectation μ and standard deviation $\sigma > 0$, with the decreasing reversed failure rate (DRFR). We are interested in finding the bounds on expectations of such standardized order statistic, i.e.

$$E\left(\frac{X_{j:n} - \mu}{\sigma}\right),$$

$j = 1, \ldots, n$. In order to determine the bounds, we need to find the projections of functions depending on the densities of the considered order statistic onto a particular convex cone of functions which are nondecreasing and convex. In this case the results of [1] are helpful and allow to obtain the desired bounds.

References

1. Danielak, K. (2003). Sharp upper mean-variance bounds for trimmed means from restricted families. *Statistics 37*, 305–324.

Goodness-of-fit tests for imprecise data

Przemysław Grzegorzewski[1,2], **Oliwia Gadomska**[1]

[1] Warsaw University of Technology
[2] Polish Academy of Sciences, Warsaw

Abstract

The problem of testing goodness-of-fit for $k \geq 2$ distributions based on imprecise data is considered. New permutation tests for fuzzy random samples are proposed. The first one is based on distances between fuzzy sample means (see [1]). The idea of the second test goes back to the k-NN classification method, which assigns an object to a class most common among its k nearest neighbors. It can be shown that by an appropriate construction of a test statistic which counts the number of nearest neighbors between and within samples it is possible to check whether available fuzzy samples come or not from the same distribution (see [2]). Finally, the goodness-of-fit test based on the so-called energy distance (see [3]) is also considered.
All suggested testing procedures are completely distribution-free which is of extreme importance in statistical reasoning with fuzzy data. Besides deriving the aforementioned tests we also compare their basic statistical properties and discuss a case study related to quality assessment illustrating the suggested procedures.

Keywords

Energy distance, Fuzzy data, Goodness-of-fit test, k-nearest neighbor, Permutation test, Random fuzzy number.

References

1. Grzegorzewski, P. (2020). Permutation k-sample goodness-of-fit test for fuzzy data. In: *Proceedings of the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2020).*
2. Grzegorzewski, P. and O. Gadomska (2021). Nearest neighbor tests for fuzzy data. In: *Proceedings of the 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2021).*
3. Székely, G.J. and M.L. Rizzo (2013). Energy statistics: A class of statistics based on distances, *J. Statist. Plann. Inference 8*, 1249–1272.

Estimation of block structured covariance matrix

Malwina Janiszewska[1], Augustyn Markiewicz[1], Monika Mokrzycka[2]

[1] Poznań University of Life Sciences
[2] Polish Academy of Sciences, Poznań

Abstract

We consider estimation of the covariance matrix with some specific block structure. The structured maximum likelihood estimator cannot be obtained explicitly and must rely on an iteration procedure. Therefore we propose alternative estimators obtained by minimization of the Frobenius norm. The estimator has a simple explicit formula that is easy to compute. However, this estimator is not always positive definite and in such a case we propose a method of its improvement. Estimators derived using both methods are compared.

Keywords

Block covariance structure, Estimation, Frobenius norm, Maximum likelihood estimation.

Conditional reliability properties of k-out-of-n system composed of different types of components

Krzysztof Jasiński

Nicolaus Copernicus University in Toruń

Abstract

We study reliability properties of k-out-of-n system consisting of l $(1 \leq l \leq n)$ different types of components with discrete, independent lifetimes. We obtain some conditional survival functions of lifetime of a used system. Next, we use it to calculate two conditional failure probabilities of k-out-of-n systems and show that they are equal to unconditional failure probability of a k-out-of-$(n-r)$ system, $r < n-k+1$. These results are extended versions of the respective ones existing in the literature.
The results are published in [1].

Keywords

Discrete lifetime distributions, k-out-of-n system, Not identically distributed random variables, Reliability theory.

References

1. Jasiński, K. (2021). Some conditional reliability properties of k-out-of-n system composed of different types of components with discrete independent lifetimes. *Metrika 84*, 1241–1251.

Various tests for independence under block compound symmetry covariance structure

Katarzyna Filipiak[1], Mateusz John[1], Daniel Klein[2]

[1] Poznań University of Technology
[2] P. J. Šafárik University, Košice, Slovakia

Abstract

The goal of this talk is to verify the hypothesis related to independence of features between any two time points under the doubly multivariate model with block compound symmetry covariance structure. The Rao score test is determined for such a hypothesis and then it is compared with the test based on the likelihood ratio, F-test and Roy's largest root test with respect to their powers and asymptotic convergence of their distributions.
Finally, Rao score test is applied to real data example.

Keywords

Testing, Rao score test, Likelihood ratio test, F-test, Roy's test.

Acknowledgements

The author was supported by the project no. 0213/SIGR/2154.

Variable selection in linear model via generalized SLOPE

Michal Kos

University of Wrocław

Abstract

We introduce a new estimator for the vector of coefficients in the linear model $y = X\beta + \epsilon$, where $y \in R^n$ – a response vector, $X \in M_{n\times p}$ – a design matrix, $\beta^* \in R^p$ - an unknown, true parameters vector, $\epsilon \in R^p$ – an error vector such that $E(\epsilon) = 0$ and $cov(\epsilon) = \sigma^2 Id$.
The Generalized SLOPE (GS) estimator is defined in a following way:

$$\hat{\beta}^{GS}(y, X, \lambda, U) = arg\min_b \left(0.5\|y - Xb\|_2^2 + 0.5\|Ub\|_2^2 + \sum_{i=1}^{p} \lambda_i |b|_{\tau(i)}\right)$$

where $U \in M_{u\times p}$ is a regularization matrix, $\lambda = (\lambda_1, ..., \lambda_p)$ is a positive nonincreasing sequence, $|b| = (|b_1|, ..., |b_p|)'$ and for any vector $w \in R^p$, τ is a permutation of a set $\{1, ..., p\}$ such that $w_{\tau(1)} \geq w_{\tau(2)} \geq .... \geq w_{\tau(p)}$.
This procedure is a generalization of SLOPE and ELASTIC NET procedures. During the session we shall present new results illustrating that GS with properly chosen regularization parameters λ and U, controls FDR at level q, when explanatory variables are correlated. This results generalize theorem 1.1 presented in [1], which proves that SLOPE with properly chosen λ, controls FDR at level q, for orthogonal design matrices ($X'X = \mathbf{I}$).

Keywords

Generalized SLOPE, Variable Selection, False Discovery Rate, Linear Model.

References

1. Bogdan, M., van den Berg, E., Sabatti, C., Su, W. and E.J. Candes (2015). SLOPE – adaptive variable selection via convex optimization. *Ann. Appl. Stat. 9*, 1103–1140.
2. Kos, M. and M. Bogdan (2020). On the asymptotic properties of SLOPE. *Sankhya A 82*, 499–532.

Adaptive MCMC for multimodal targets

Chris Holmes[1], Krzysztof Łatuszyński[2], Emilia Pompe[1]

[1] University of Oxford, UK
[2] University of Warwick, UK

Abstract

We propose a new Monte Carlo method for sampling from multimodal distributions [1]. The idea of this technique is based on splitting the task into two: finding the modes of a target distribution π and sampling, given the knowledge of the locations of the modes. The sampling algorithm relies on steps of two types: local ones, presering the mode; and jumps to regions associated with different modes. Besides, the method learns the optimal parameters of the algorithm while it runs, without requiring user intervention. Our technique should be considered as a flexible framework, in which the design of moves can follow various strategies known from the broad MCMC literature. In order to design an adaptive scheme that facilitates both local and jump moves, we introduce an auxiliary variable representing each mode and we define a new target distribution $\tilde{\pi}$ on an augmented state space $\mathcal{X} \times \mathcal{I}$, where $\mathcal{X}$ is the original state space of π and $\mathcal{I}$ is the set of the modes. As the algorithm runs and updates its parameters, the target distribution $\tilde{\pi}$ also keeps being modified. This motivates a new class of algorithms, Auxiliary Variable Adaptive MCMC. We prove general ergodic results for the whole class before specialising to the case of our algorithm.

Keywords

Adaptive MCMC, Ergodicity, Multimodality.

References

1. Pompe, E., Holmes, C. and K. Łatuszyński (2020). A framework for adaptive MCMC targeting multimodal distributions. *Ann. Statist. 48*, 2930–2952.

Multiple testing of conditional independence hypotheses using information-theoretic measures

Małgorzata Łazęcka[1,2], Jan Mielniczuk[1,2]

[1] Polish Academy of Sciences, Warsaw
[2] Warsaw University of Technology

Abstract

We study the multiple testing problem for which individual hypotheses of interest correspond to conditional independence of the two variables X and Y given each of the several conditioning variables. Formally, the problem can be stated as testing p individual hypotheses

$$H_{0,i}: \ X \perp Y \mid Z_i$$

where $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and $Z_i \in \mathcal{Z}_i$ are some observed discrete random variables for $i = 1, \dots, p$. Our aim is to construct a test which controls type I error under global null $H_0 = \cap_{i=1}^{p} H_{0,i}$ when all null hypotheses $H_{0,i}$ are true. We introduce a direct approach based on Joint Mutual Information (JMI) statistics which restates the problem as a problem of single hypothesis testing. In this talk I will present the distribution of the test statistics JMI and show that it can be well numerically approximated for a single data sample. I will also present the performance of the corresponding test on artificial data sets.

Keywords

Conditional independence, Joint mutual information, Multiple testing, Weighted chi-square distribution.

Linear sufficiency and admissibility under various loss functions

Augustyn Markiewicz

Poznań University of Life Sciences

Abstract

The problem of admissible linear estimation under the Gauss-Markov model was extensively studied by Rao in [7]. Rao's results were developed in [3] and [2] under a possibly singular model. However, the linear sufficiency was studied and characterized in [1] and [4]. The subject of linear sufficiency and admissibility was considered in [5] and [6].
The purpose of this paper is to present estimators that satisfy both of these properties, along with some new supplementary results.

Keywords

Admissibility, Linear sufficiency.

References

1. Baksalary, J.K. and R. Kala (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss-Markoff model. *Ann. Statist. 9*, 913–916.
2. Baksalary, J.K., Markiewicz, A. and C.R. Rao (1995). Admissible linear estimation in the general Gauss-Markov model with respect to an arbitrary quadratic risk function. *J. Statist. Plann. Inference 44*, 341–347.
3. Baksalary, J.K., Rao, C.R. and A. Markiewicz (1992). A study of the "natural restrictions" on the estimation problems in the singular Gauss-Markov model. *J. Statist. Plann. Inference 31*, 335–351.
4. Drygas, H. (1983). Sufficiency and completeness in the general Gauss-Markov model. *Sankhya A 45*, 88–98.
5. Markiewicz, A. (1996). Characterization of general ridge estimators. *Stat. Probab. Lett 27*, 145–148.
6. Markiewicz, A. and S. Puntanen (2009). Admissibility and linear sufficiency in linear model with nuisance parameters. *Statist. Papers 50*, 847–854.
7. Rao, C.R. (1976). Estimation of parameters in a linear model. *Ann. Statist. 4*, 1023–1037.

The comparison of the estimators of linearly structured covariance matrix

Adam Mieldzioc

Poznań University of Life Sciences

Abstract

The aim of the work is to derive a well-conditioned estimator of the covariance matrix with some linear structure. In this case, the determination of maximum likelihood estimator can be challenging, because in general does not exist its explicit formula. The procedure of its derivation relies on iteration algorithm and might be time-consuming especially in the high-dimensional case, where many characteristics are observed.
To ensure the linear structure of the estimator, the sample covariance matrix from the model without any restrictions is projected on the proper space. The projection can be indefinite ([1]). Therefore, we adopt *shrinkage method, cf. [2,3]* to obtain positive definite estimator.
The shrinkage estimator is compared with the maximum likelihood estimator using the selected loss functions.

Keywords

Banded Toeplitz matrix, Shrinkage, MLE, Frobenius norm

References

1. Filipiak K., Markiewicz A., Mieldzioc A., Sawikowska A. (2018): On projection of a positive definite matrix on a cone of nonnegative definite Toeplitz matrices. *Electronic Journal of Linear Algebra 33*, 74-?82.
2. John, M. and Mieldzioc, A. (2020). The comparison of the estimators of banded Toeplitz covariance structure under the high-dimensional multivariate model. *Communications in Statistics - Simulation and Computation 49(3)*, 734–752.
3. Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis 88(2)*, 365–411.

Joint estimation of posterior probability and propensity score for positive and unlabelled data

Konrad Furmańczyk[1], Jan Mielniczuk[2,3], Paweł Teisseyre[3], Wojciech Rejchel[4]

[1] Warsaw University of Life Sciences
[2] Warsaw University of Technology
[3] Polish Academy of Sciences, Warsaw
[4] Nocolaus Copernicus University in Toruń

Abstract

Let X be a random variable corresponding to a feature vector, $Y \in \{0,1\}$ be a true class label and $S \in \{0,1\}$ an indicator of an example being labelled ($S = 1$) or not ($S = 0$). We assume that there is some unknown distribution $P_{Y,X,S}$ such that $(Y_i, X_i, S_i), i = 1, \ldots, n$ is iid sample drawn from it. Observed data consists of $(X_i, S_i), i = 1, \ldots, n$ (single sample scenario [1]). Only positive examples ($Y = 1$) can be labelled, i.e. $P(S = 1|X, Y = 0) = 0$. Thus we know that $Y = 1$ when $S = 1$ but when $S = 0$, Y can be either 1 or 0. Our primary aim is to learn binary posterior distribution of Y given $X = x$ i.e. $y(x) = P(Y = 1|X = x)$ when we only observe samples from distribution of (X, S), where $S = Y$ with a certain probability.
To this end we define binary posterior distribution of S given $X = x$ equal $s(x) = P(S = 1|X = x)$ and propensity score function $e(x) = P(S = 1|Y = 1, X = x)$. We note that by conditioning on Y we have

$$s(x) = P(Y = 1|X = x)P(S = 1|Y = 1, X = x) = y(x)e(x). \qquad (1)$$

The significant limitation of almost all existing methods lies in assuming that the propensity score function is constant (SCAR assumption), which is unrealistic in many practical situations. Avoiding this assumption, we consider parametric approach to the problem of joint estimation of posterior probability and propensity score functions (see [2]). We show that under mild assumptions when both functions have the same parametric form (e.g. logistic with different parameters) the corresponding parameters are identifiable up to a swap. Motivated by this, we propose two approaches to their estimation: joint maximum likelihood method and the second approach based on alternating maximization of two Fisher consistent expressions. Our experimental results show that the proposed methods are comparable or better than the Expectation-Maximisation algorithm considered in this context.

Keywords

Logistic regression, Positive and unlabelled data, Empirical risk minimisation, Propensity score function.

References

1. Bekker, J. and J. Davis (2020). Learning from positive and unlabelled data: a survey. *Machine Learning 109*, 739–768.
2. Furmańczyk, K., Mielniczuk, J., Teisseyre, P. and W. Rejchel. Joint estimation of posterior probability and propensity score for positive and unlabelled data. Submitted.

Discrepancy measures in the context of power study for testing covariance structures

Katarzyna Filipiak[1], Daniel Klein[2], Monika Mokrzycka[3]

[1] Poznań University of Technology
[2] P. J. Šafárik University, Košice, Slovakia
[3] Polish Academy of Sciences, Poznań

Abstract

To compare the powers of the tests, the measure of discrepancy between null and alternative hypotheses is required. One method of measuring such discrepancy is based on one parameter, we called it naive method, while more complex methods are based on minimization of some discrepancy functions over the set of considered structures. In this talk naive method, Frobenius norm, as well as entropy and quadratic loss functions will be used as discrepancies between hypotheses and compared with respect to the power of the likelihood ratio and Rao score tests.

Keywords

Separable structure, Discrepancy, Frobenius norm, Entropy loss function, Quadratic loss function.

Sensitive questions in survey statistics: some new versions of Item Count Technique

Wojciech Niemiro

[1] University of Warsaw
[2] Nicolaus Copernicus University in Toruń

Abstract

The item count technique (ICT) is a survey methodology that is designed to elicit respondents' truthful answers to sensitive questions such as racial prejudice and drug use [3]. Respondents are randomly assigned either to the treatment group or to the control group. Those in the control group are asked only some neutral (masking) question. Respondents in the treatment group are asked the same neutral question *and* the sensitive question of interest. They report the sum of answers.
Standard way of analysing ICT data is to use simple moment estimates. Imai [2] proposed a linear regression model and maximum likelihood computated via the EM algorithm.
In my talk I will describe some recent modifications and improvements of ICT. First, in article [1] we proposed using *two* treatment groups: respondents report either the sum of answers or the difference, depending on the group they are assigned to. Second, we apply EM a nonparametric discrete model to increase efficiency and robustness. Third, our approach works fine with the logistic regression model.
The talk will report the results obtained together with Barbara Kowalczyk and Robert Wieczorkowski.

Keywords

Hidden variable, Maximum Likelihood, Expectation-Maximization, Questionnaires, Logistic regression.

References

1. Kowalczyk, B., Niemiro, W. and R. Wieczorkowski (2021). Item count technique with a continuous or count control variable for analyzing sensitive questions in surveys. To appear in *Journal of Survey Statistics and Methodology*.
2. Imai, K. (2011). Multivariate regression analysis for the Item Count Technique. *J. Amer. Statist. Assoc. 106*, 494.
3. Chaudhuri, A. and T.C. Christofides (2007). Item Count Technique in estimating the proportion of people with a sensitive feature. *J. Statist. Plann. Inference 137*, 589-593.

On the choice of the optimal linear combination of two order statistics in quantile estimation distributions with given moments

Luiza Pańczyk, **Mariusz Bieniek**

Maria Curie–Skłodowska University, Lublin

Abstract

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ denote the order statistics of the sample of size n. We consider the problem of the estimation of population quantile $F^{-1}(p)$ of order $p \in (0,1)$ by the use of appropriately chosen linear combination of two adjacent order statistics $(1-\alpha)X_{j:n} + \alpha X_{j+1:n}$. For a given sample size and quantile order, we are looking for such values j and α, for which the quantile estimator of the above form is optimal in selected classes of distributions with given moments. In last year's talk we considered distributions with finite variance and we showed that optimal choice α does not coincide with the traditional choice $\alpha = np$, that is, linear interpolation of the quantile function. This time we are considering two classes of particular distributions: the wider class with finite mean and the narrower one — of the distributions suppoerted on a bounded interval. We will show that the choice in these classes for most values of the quantile order $p \in (0,1)$ coincides with the previously obtained choice.

Estimation of parameters and quantiles of the Weibull distribution

Alicja Jokiel-Rokita[1], <u>Sylwester Piątek</u>

[1]Wrocław University of Science and Technology

Abstract

The Weibull distribution is used in various branches of statistics, such as reliability, risks analysis, and quality control. It is also a suitable model for wind speed frequency distribution and it has been widely used in many fields of the wind energy industry ([2]).

In the presentation, we make some review of estimation methods of the Weibull distribution parameters. We also present three new estimators which lead to three new plug-in estimators of quantiles. One of them is a slight modification of the maximum likelihood estimators (MLE's) and two of them are based on nonparametric estimators of the Gini coefficient. We compare the small sample performance (in terms of bias and mean squared error) of the known and new estimators and extreme quantiles. Similar comparison, but only for estimators of parameters, was performed in [1] and [3]. We obtain, among others, that for the shape parameter greater than one, the proposed modification of MLE is better or as good as known methods in the estimation of parameters and extreme quantiles.

Keywords

Extreme quantiles, Gini coefficient, maximum likelihood estimation, Monte Carlo simulation.

References

1. Gebizlioglu, O.L., Şenoğlu, B. and Y.M. Kantar (2011). Comparison of certain value-at-risk estimation methods for the two-parameter Weibull loss distribution. *J. Comput. Appl. Math. 235*, 3304–3314.
2. Kang, D., Ko, K. and J. Huh (2018). Comparative study of different methods for estimating Weibull parameters: A case study an Jeju Islands, South Korea. *Energies 11*, 356.
3. Teimouri, M., Hoseini, S.M. and S. Nadarajah (2013). Comparison of estimation methods for the Weibull distribution. *Statistics 47*, 93–109.

Objective Bayes factors in linear models

Piotr Pokarowski

University of Warsaw

Abstract

Bayes factors are the basic indicators for assessing the incompatibility of data with the null hypothesis with the alternative, and therefore they compete with p-values. The statistical literature already presents at least a dozen objective Bayes factors for inference in linear models and there is no consensus as to which factor is the best. The purpose of the presentation is to compare whether there are significant differences between these factors - both from a theoretical and practical point of view.

Selection consistency for high-dimensional categorical data

Szymon Nowakowski[1], Piotr Pokarowski[1], Wojciech Rejchel[2]

[1] University of Warsaw
[2] Nicolaus Copernicus University in Toruń

Abstract

Sparse prediction with categorical data is challenging even for a moderate number of variables, because one parameter is roughly needed to encode one category or level. The group lasso is a well known and efficient algorithm for selection continuous or categorical variables, but all estimates related to a selected factor usually differ, so a fitted model may not be sparse. To make a group lasso solution sparse, we propose to merge levels of the selected factor, if a difference between its corresponding estimates is less than some predetermined threshold. We prove that under weak conditions our algorithm recovers the true, sparse linear or logistic model even for the high-dimensional scenario, that is when a number of parameters is greater than a sample size. To our knowledge, selection consistency has been proven many times for different algorithms fitting sparse models with categorical variables, but our result is the first for the high-dimensional scenario.

Shrinkage estimators for uplift regression

Krzysztof Rudaś[1,2], Magdalena Grabarczyk[1]

[1] Warsaw University of Technology
[2] Polish Academy of Sciences, Warsaw

Abstract

To evaluate efficiency of marketing campaign or medical treatment, we divide our population into two subgroups: treatment (on which action is taken) and control (on which no action is taken) and predict the difference between effects observed in both groups.
In [1] were presented two typical methods of estimating this difference called the double estimator and the uplift estimator. In [2] were presented shrinkage methods for the double approach which have smaller MSE than original estimator. However the double method does not work good for cases where sizes of treatment and control groups are similar to number of features. In this case the uplift estimator is better approach. In our presentation we propose new shrinkage methods for the uplift estimator, basing on James-Stein and Ohtani approach for the linear regression. Statistical properties of this estimators are studied.

Keywords

Uplift modeling, Linear regression, Shrinkage estimators.

References

1. Rudaś, K. and S. Jaroszewicz (2018). Linear regression for uplift modeling. *Data Min. Knowl. Discov. 32*, 1275–1305.
2. Rudaś, K. and S. Jaroszewicz (2020). Shrinkage estimators for uplift regression. *Machine Learning and Knowledge Discovery in Databases*, 607–623.

Bounds on the expectations of L-statistics from iid symmetric samples in various scale units

Tomasz Rychlik

Polish Academy of Sciences, Warsaw

Abstract

We consider the order statistics $X_{1:n}, \dots, X_{n:n}$ based on independent identically symmetrically distributed random variables. We determine sharp upper bounds in the properly centered linear combinations of order statistics $\sum_{i=1}^{n} c_i(X_{i:n} - \mu)$, where $(c_1, \dots, c_n)$ is an arbitrary vectors of coefficients from the n-dimensional real space, and μ is the symmetry center of the parent distribution, in various scale units. The scale units are constructed on the basis of absolute central moments of the parent distribution of various orders. The bounds are specified for single order statistics. The lower bounds are immediately concluded from the upper ones.

Keywords

Independent identically distributed random variables, Symmetric distribution, Order statistic, L-statistic, Sharp bound.

Model recovery by SLOPE

Tomasz Skalski[1,2], Małgorzata Bogdan[3], Piotr Graczyk[2], Bartosz Kołodziejek[4], Patrick Tardivel[5], Maciej Wilczyński[1]

[1] Wrocław University of Science and Technology
[2] University of Angers, France
[3] University of Wrocław
[4] Warsaw University of Technology
[5] University of Burgundy, France

Abstract

The Sorted L-One Penalized Estimation (SLOPE), a generalization of the LASSO estimator, was introduced by Bogdan, van den Berg, Sabatti, Su and Candès in 2015. In addition to maintaining the properties of LASSO, SLOPE is useful in recovery of blocks of equal absolute values of coordinates, which can be seen as a property leading to further reduction of dimension.
The most important result of our research is the equivalence of the Model Recovery by SLOPE with probability near 1, when the signal to noise ratio is large, with any of two equivalent irrepresentability conditions: with the analytic condition, generalizing the irrepresentability condition for LASSO given by Zou in 2006 and with the geometric condition.
Our results are true both in high-dimensional setting ($p > n$) and in low-dimensional ($p \leq n$).

Acknowledgements

This research was supported by a French Government Scholarship.

Permutation test for the multivariate coefficient of variation in factorial designs

Marc Ditzhaus[1], Łukasz Smaga[2]

[1] TU Dortmund University
[2] Adam Mickiewicz University, Poznań

Abstract

New inference methods for the multivariate coefficient of variation are presented. While there are various testing procedures for this parameter in the univariate case, it is less known how to do inference in the multivariate setting appropriately. There are some existing procedures but they rely on restrictive assumptions on the underlying distributions. We tackle this problem by applying Wald-type statistic in the context of general, potentially heteroscedastic factorial designs. The resulting procedure is shown to be asymptotically valid under the null hypothesis and consistent under general alternatives. To improve the finite sample performance, we suggest a permutation version of the test and show that the test's asymptotic properties can be transferred to it. An exhaustive simulation study compares the new tests and existing methods.

Keywords

Coefficient of variation, General factorial designs, Hypothesis testing, Multivariate analysis, Permutation method.

References

1. Aerts, S. and G. Haesbroeck (2017). Robust asymptotic tests for the equality of multivariate coefficients of variation. *Test 26*, 163–187.
2. Albert, A. and L. Zhang (2010). A novel definition of the multivariate coefficient of variation. *Biom. J. 52*, 667–675.
3. Pauly, M. and Ł. Smaga (2020). Asymptotic permutation tests for coefficients of variation and standardized means in general one-way ANOVA models. *Stat. Methods Med. Res. 29*, 2733–2748.
4. Voinov, V. and M. Nikulin (1996). *Unbiased Estimators and Their Applications. Vol. 2, Multivariate Case.* Kluwer, Dordrecht.

The Weibull lifetime model with randomized failure-free time

Piotr Sulewski[1], Magdalena Szymkowiak[2]

[1] Pomeranian University in Słupsk
[2] Poznań University of Technology

Abstract

The paper shows that treating failure-free time in the three-parameter Weibull distribution - not as the constant, but as the random variable - makes the resulting distribution much more flexible for the price of only one parameter more.

Keywords

Weibull lifetime model, Randomized failure-free time, Compound Weibull distributions.

Acknowledgements

The authors thank professor Antoni Drapella, former Director of Institute of Mathematics at Pomeranian University in Słupsk, for providing assistance in formulating basics of the method presented.

The second author was partially supported by Poznań University of Technology under grant 0211/ SBAD/0121.

References

1. Drapella, A. (1993). Complementary weibull distribution: Unknown or just forgotten. *Quality and Reliability Engineering International 9*, 383–385.
2. Drapella, A. (1999). An improved failure-free time estimation method. *Quality and Reliability Engineering International 15*, 235–238.
3. Kendall, M. and A. Stuart (1961). *The advanced theory of statistics, Vol. 2.* Charles Griffin and Company.
4. O'Connor, P. (1985). *Practical reliability engineering.* Fifth Edition. Willey.
5. Szymkowiak, M. (2018). Characterizations of distributions through aging intensity. *IEEE Transactions on Reliability 67*, 446–458.
6. Szymkowiak, M. (2020). *Lifetime analysis by aging intensity function.* Springer.

Two methods of conjoint summands of generating bivariate and trivariate normal pseudo-random numbers

Piotr Sulewski

Pomeranian University in Słupsk

Abstract

The primary aim of the paper is to put forward methods of generating two dimensional (2D) and three dimensional (3D) normal pseudo-random numbers (NPRNs). The secondary aim of the paper is to compare the existing methods of generating 2D and 3D NPRNs with new proposals using goodness-of-fit tests (GoFTs), generation time and specially defined measures. The obtained results indicate that the methods based on uniform $(0, 1)$ random variables are faster than the others (have the best computational complexity) and also stand out from the other measures and GoFTs used.

Keywords

Bivariate normal samples, Trivariate normal samples, Multivariate gaussian samples, Multivariate normal generators, Monte Carlo simulation.

Acknowledgements

The author thanks professor Antoni Drapella, former Director of Institute of Mathematics at Pomeranian University in Słupsk, for providing assistance in formulating basics of the method presented.

References

1. Ahrens, J.H. and U. Dieter (1972). Computer methods for sampling from the exponential and normal distributions. *Communication of the ACM 15*, 873–882.
2. Barr, D.R. and N.L. Slezak (1972). A comparison of multivariate normal generators. *Communications of the ACM 15*, 1048–1049.
3. Cheng, R.C.H. (1985). Generation of multivariate normal samples with given sample mean and covariance matrix. *J. Stat. Comput. Simul. 21*, 39–49.
4. Doornik, J.A. and H. Hansen (2008). An Omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics 70*, 927–939.
5. Henze, N. and B.A. Zirkler (1990). Class of Invariant Consistent Tests for Multivariate Normality. *Comm. Statist. Theory Methods 19*, 3595–3618.

6. Mardia, K.V. (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhya A 36*, 115–128.
7. Mi, X., Miwa, T. and T. Hothorn (2009). mvtnorm: New numerical algorithm for multivariate normal probabilities. *The R Journal 1*, 37–39.
8. Thomas, D.B. and W. Luk (2008). Multivariate Gaussian random number generation targeting reconfigurable hardware. *ACM Transactions on Reconfigurable Technology and Systems 1*, 1–29.
9. Villasenor Alva, J.A. and E.G. Estrada (2009). A generalization of Shapiro-Wilk's test for multivariate normality. *Comm. Statist. Theory Methods 38*, 870–1883.

Testing of factor effects in a three level model

Anna Szczepańska-Álvarez

Poznań University of Life Sciences

Abstract

In this talk a three level model is considered, with a covariance structure $\Sigma \otimes \Psi^1 \otimes \Psi^2$, where Σ is an arbitrary positive definite matrix, and Ψ^1 and Ψ^2 both are correlation matrices. With this choice of matrices there is a unique parametrization and all parameters are estimable. Moreover, we assume that matrices Ψ^1 and Ψ^2 have compound symmetric structure. The maximum likelihood estimators are determined using the spectral decomposition $\Psi^1 = \Gamma_1 \Lambda_1 \Gamma_1'$, $\Psi^2 = \Gamma_2 \Lambda_2 \Gamma_2'$. We are interested to test the factor effects and we formulate two hypotheses: $H_0^1 : \Lambda_1 = I_p$ and $H_0^2 : \Lambda_1 = I_p, \Lambda_2 = I_q$. To verify the null hypothesis we use Rao's score statistic for testing a subset of parameters. The results are illustrated by data from a fermentation process.

Adaptivity via a discrepancy principle for Poisson inverse problems

Zbigniew Szkutnik

AGH University of Science and Technology, Cracow

Abstract

Let $\mathcal{K} : H_1 \to H_2$ be an operator between some Hilbert spaces H_1, H_2. The general stochastic inverse problem consists in estimating $f \in H_1$, given some stochastic data with distribution depending on $g \in H_2$ and with $g = \mathcal{K}f$. Let $\hat{f}_\alpha$ be an estimator of f obtained with some regularization parameter α. Data-driven procedures of choosing α, preferably leading to estimators that are asymptotically optimally adaptive over smoothness classes, are of central interest.
In the best studied special case, g is observed as $\tilde{g}$ in Gaussian white noise W in H_2, i.e. $\tilde{g} = \mathcal{K}f + \delta W$, with some known noise level $\delta \to 0$. In Poisson inverse problems, the observed object is a Poisson process with an intensity function ng and with known $n \to \infty$.
Discrepancy principle (DP) chooses the value of α for which discrepancy $\|\mathcal{K}\hat{f}_\alpha - \tilde{g}\|$ matches a ńaturalłevel. For non-stochastic inverse problems, it produces adaptive rate optimal solutions. In the white noise model, since W is not an H_2-valued process, one often works with the pre-conditioned model, which leads to smoothed discrepancy $\|\mathcal{K}^*(\mathcal{K}\hat{f}_\alpha - \tilde{g})\| = \|(\mathcal{K}\mathcal{K}^*)^{1/2}(\mathcal{K}\hat{f}_\alpha - \tilde{g})\|$. It turns out, however, that the resulting estimates are, in some typical setups, rate suboptimal, essentially because the residuals are smoothed too much. More flexible variant of smoothed discrepancy of the form $\|(\mathcal{K}\mathcal{K}^*)^{\gamma/2}(\mathcal{K}\hat{f}_\alpha - \tilde{g})\|$, with some $\gamma \geq 0$ was proposed by Stankewitz (2020). For truncated SVD estimators, he proved (in a somewhat different discretized setup) that "moderate smoothing of the residuals can be used to adapt over classes of signals with varying smoothness, while oversmoothing yields suboptimal convergence rates".
In this talk, a special version of DP, suitable for Poisson inverse problems, will be discussed. As in the white noise case, it proves convenient to work with the pre-conditioned model. The value of α is selected to satisfy

$$\|\mathcal{K}^*\mathcal{K}\hat{f}_\alpha - \hat{q}\|^2 = \tau\hat{\delta}^2$$

with some fixed, positive τ and with some specific, unbiased estimators $\hat{q}(\cdot)$ and $\hat{\delta}^2$ of, respectively, $q(\cdot) := (\mathcal{K}^*g)(\cdot)$ and the ńoise level"$\delta^2 := \mathrm{E}\|\hat{q} - q\|^2$. Although spectral filter estimators combined with this version of DP adapt to the unknown smoothness, defined via source conditions or Sobolev-type

ellipsoids, the resulting convergence rates are, as in the white noise case, suboptimal. Changing the degree of residual smoothing seems, however, to open encouraging perspectives also in the Poissonian case.

Conditions on system signatures assuring the preservation of some distribution properties of component lifetimes by system lifetime

Barry C. Arnold[1], Tomasz Rychlik[2], Magdalena Szymkowiak[3]

[1] University of California, Riverside, USA
[2] Polish Academy of Sciences, Warsaw
[3] Poznań University of Technology

Abstract

We discuss the properties of the lifetime distribution of a system with components whose lifetimes are identically distributed, and their joint distribution admits the Samaniego signature representation. We present the sufficient conditions on the system signatures under which the monotone failure rate of the components lifetimes is inherited by the system lifetime distribution.

Keywords

Components and system lifetimes, Samaniego signatures, Monotone failure rate.

Acknowledgements

This research is partially supported by PUT under grant 0211/SBAD/0121.

References

1. Arnold, B.C., Rychlik, T. and M. Szymkowiak (2021). Preservation of distributional properties of component lifetimes by system lifetimes. Submitted.
2. Navarro, J., Rychlik, T. and F. Spizzichino (2021). Conditions on marginals and copula of component lifetimes for signature representation of system lifetime. *Fuzzy Sets and Systems 415*, 99–117.
3. Rychlik, T. and M. Szymkowiak (2021). Signature conditions for distributional properties of system lifetimes if component lifetimes are iid exponential. *IEEE Transactions on Reliability*, DOI: 10.1109/TR.2021.3119463.
4. Samaniego, F.J. (1985). On closure of the IFR class under formation of coherent systems. *IEEE Transactions on Reliability 34*, 69–72.

Feature selection for positive-unlabelled multi-label classification

Paweł Teisseyre

[1] Polish Academy of Sciences, Warsaw
[2] Warsaw University of Technology

Abstract

We consider the positive-unlabelled multi-label scenario [1,2] in which multiple target variables are not observed directly. Instead, we observe surrogate variables indicating whether the target variables are labelled or not. The presence of the label means that the variable is positive. The absence of the label means that the variable can be either positive or negative. We analyse embedded feature selection methods based on two weighted penalized empirical risk minimization frameworks. In both proposed methods we use $\ell_{2,1}$ penalty which allows to simultaneously select relevant features in the models corresponding to the considered target variables. In the experiments we analyse the predictive power of the considered methods for different labelling schemes.

References

1. Bekker, J. and J. Davis (2020). Learning from positive and unlabelled data: a survey. *Machine Learning 109*, 719–760.
2. Teisseyre, P. (2021). Classifier chains for positive unlabelled multi-label learning. *Knowledge-Based Systems 213*, 1–16.

A nonparametric test for paired data

Grzegorz Wyłupek

University of Wrocław

Abstract

The paper proposes a weighted Kolmogorov-Smirnov type test for the two-sample problem when the data is paired. The asymptotic distribution of the test statistic under the null model is derived. The dependence of both finite sample and asymptotic distribution of the test statistic from the dependence structure of the data forces the usage of the wild bootstrap technique for the inference. The related wild bootstrap test turns out to be a consistent asymptotically level α test. An application of the finite sample correction allows the test to keep the level well. An extensive simulation study demonstrates good finite sample behaviour of the test in comparison to the existing procedures.

References

1. Ghattas, B., Pommeret, D., Reboul, L. and A.F. Yao (2011). Data driven smooth test for paired populations. *J. Statist. Plann. Inference 141*, 262–275.
2. Martínez-Camblor, P. (2010). Nonparametric k-sample test based on kernel density estimator for paired design. *Comput. Statist. Data Anal. 54*, 2035–2045.
3. Nelsen, R.B. (2006). *An Introduction to Copulas.* Springer, New York.
4. Pereira, L.A., Taylor-Rodríguez, D. and L. Gutiérrez (2020). A Bayesian nonparametric testing procedure for paired samples. *Biometrics 76*, 1133–1146.
5. Quessy, J-F. and F. Éthier (2012). Cramér-von Mises and characteristic function tests for the two and k-sample problems with dependent data. *Comput. Statist. Data Anal. 56*, 2097–2111.
6. Vexler, A., Gurevich, G. and A.D. Hutson (2013). An exact density-based empirical likelihood ratio test for paired data. *J. Statist. Plann. Inference 143*, 334–345.
7. Wyłupek, G. (2021). A nonparametric test for paired data. Submitted.

Część V

Uczestnicy

1. **Karol Andrzejczak**
Instytut Matematyki
Politechnika Poznańska
karol.andrzejczak@put.poznan.pl
2. **Michał Boczek**
Instytut Matematyki
Politechnika Łódzka
michal.boczek.1@p.lodz.pl
3. **Krzysztof Bogdan**
Wydział Matematyki
Politechnika Wrocławska
krzysztof.bogdan@pwr.edu.pl
4. **Małgorzata Bogdan**
Instytut Matematyki
Uniwersytet Wrocławski
malgorzata.bogdan@uwr.edu.pl
5. **Jarosław Brzeziński**
Polsko-Japońska Akademia Technik Komputerowych w Warszawie
jaroslaw.brzezinski@pj.edu.pl
6. **Anna Dudek**
Wydział Matematyki Stosowanej
Akademia Górniczo-Hutnicza w Krakowie
aedudek@agh.edu.pl
7. **Marcin Dudziński**
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
marcin_dudzinski@sggw.edu.pl
8. **Diana Dziewa-Dawidczyk**
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
diana_dziewa_dawidczyk@sggw.pl
9. **Katarzyna Filipiak**
Instytut Matematyki
Politechnika Poznańska
katarzyna.filipiak@put.poznan.pl
10. **Paweł Filipiak**
Politechnika Poznańska
pawel.z.filipiak@student.put.poznan.pl
11. **Konrad Furmańczyk**
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
konrad_furmanczyk@sggw.edu.pl

12. **Agnieszka Geras**
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
A.Geras@mini.pw.edu.pl

13. **Bartosz Głowacki**
Akademia Górniczo-Hutnicza w Krakowie
glowacki@student.agh.edu.pl

14. **Agnieszka Goroncy**
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
gemini@mat.umk.pl

15. **Michał Gostkowski**
Katedra Ekonometrii i Statystyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
michal_gostkowski@sggw.edu.pl

16. **Przemysław Grzegorzewski**
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
Instytut Badań Systemowych
Polska Akademia Nauk, Warszawa
pgrzeg@mini.pw.edu.pl

17. **Jan Hauke**
Wydział Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej
Uniwersytet im. Adama Mickiewicza w Poznaniu
jan.hauke@amu.edu.pl

18. **Krzysztof Hernik**
Wydział Matematyki Stosowanej
Akademia Górniczo-Hutnicza w Krakowie

19. **Malwina Janiszewska**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
malwina.janiszewska@up.poznan.pl

20. **Krzysztof Jasiński**
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
krzys@mat.umk.pl

21. **Mateusz John**
Instytut Matematyki
Politechnika Poznańska
mateusz.john@put.poznan.pl

22. **Alicja Jokiel-Rokita**
Wydział Matematyki
Politechnika Wrocławska
alicja.jokiel-rokita@pwr.edu.pl

23. **Dorota Juszczak**
Instytut Ekonometrii
Szkoła Główna Handlowa w Warszawie
djuszcz@sgh.waw.pl

24. **Daniel Klein**
Instytut Matematyki
Uniwersytet Pavla Jozefa Šafárika w Koszycach, Słowacja
daniel.klein@upjs.sk

25. **Jacek Koronacki**
Instytut Podstaw Informatyki
Polska Akademia Nauk, Warszawa
jacek.koronacki@ipipan.waw.pl

26. **Michał Kos**
Instytut Matematyki
Uniwersytet Wrocławski
michal.kos@uwr.edu.pl

27. **Tomasz Kossowski**
Wydział Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej
Uniwersytet im. Adama Mickiewicza w Poznaniu
tkoss@amu.edu.pl

28. **Anastasia Koval**
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

29. **Dorota Koziol-Kaczorek**
AstraZeneca Pharma Poland
dorota.koziol-kaczorek@astrazeneca.com

30. **Patryk Krasuski**
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
patrykk@doktorant.umk.pl

31. **Mirosław Krzyśko**
Międzywydziałowy Zakład Matematyki i Statystyki
Akademia Kaliska im. Prezydenta Stanisława Wojciechowskiego
w Kaliszu
mkrzysko@amu.edu.pl

32. **Krzysztof Łatuszyński**
University of Warwick, Wielka Brytania
latuch@gmail.com

33. **Małgorzata Łazęcka**
Instytut Podstaw Informatyki
Polska Akademia Nauk, Warszawa
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
m.lazecka@mini.pw.edu.pl

34. **Ryszard Magiera**
Wydział Matematyki
Politechnika Wrocławska
ryszard.magiera@pwr.edu.pl

35. **Bartosz Majewski**
Wydział Matematyki Stosowanej
Akademia Górniczo-Hutnicza w Krakowie
bartmaje@agh.edu.pl

36. **Karolina Marek**
Wydział Matematyki Stosowanej
Akademia Górniczo-Hutnicza w Krakowie

37. **Augustyn Markiewicz**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
amark@up.poznan.pl

38. **Marek Męczarski**
Instytut Ekonometrii
Szkoła Główna Handlowa w Warszawie
mecz@sgh.waw.pl

39. **Błażej Miasojedow**
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
bmiasojedow@gmail.com

40. **Adam Mieldzioc**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
adam.mieldzioc@mail.up.poznan.pl

41. **Jan Mielniczuk**
Instytut Podstaw Informatyki
Polska Akademia Nauk, Warszawa
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
miel@ipipan.waw.pl

42. **Monika Mokrzycka**
Instytut Genetyki Roślin
Polska Akademia Nauk, Poznań
mmok@igr.poznan.pl

43. **Krzysztof Moliński**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
krzysztof.molinski@up.poznan.pl

44. **Wojciech Niemiro**
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
wniem@mimuw.edu.pl

45. **Maria Nuc**
Instytut Genetyki Roślin
Polska Akademia Nauk, Poznań
mnuc@igr.poznan.pl

46. **Anna Ogłoszka**
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

47. **Luiza Pańczyk**
Instytut Matematyki
Uniwersytet Marii Curie-Skłodowskiej w Lublinie
luiza.panczyk@o2.pl

48. **Sylwester Piątek**
sylwesterpiatek4@gmail.com

49. **Kacper Paczutkowski**
Instytut Informatyki Technicznej
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
kacper_paczutkowski@sggw.edu.pl

50. **Piotr Pokarowski**
Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski
pokar@mimuw.edu.pl

51. **Sandra Pułrolniczak**
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

52. **Wojciech Rejchel**
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
wrejchel@gmail.com

53. **Krzysztof Rudaś**
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
Instytut Podstaw Informatyki
Polska Akademia Nauk, Warszawa
krzysztof.rudas@gmail.com

54. **Leszek Rutkowski**
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

55. **Tomasz Rychlik**
Instytut Matematyczny
Polska Akademia Nauk, Warszawa
trychlik@impan.pl

56. **Aneta Sawikowska**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
aneta.sawikowska@mail.up.poznan.pl

57. **Agnieszka Siedlaczek**
Instytut Fizyki
Uniwersytet Opolski
asiedlaczek@uni.opole.pl

58. **Tomasz Skalski**
Wydział Matematyki
Politechnika Wrocławska
LAREMA
Université d'Angers, Francja
tomasz.skalski@pwr.edu.pl

59. **Łukasz Smaga**
Wydział Matematyki i Informatyki
Uniwersytet im. Adama Mickiewicza w Poznaniu
ls@amu.edu.pl

60. **Piotr Sulewski**
Instytut Nauk Ścisłych i Technicznych
Akademia Pomorska w Słupsku
piotr.sulewski@apsl.edu.pl

61. **Anna Szczepańska-Alvarez**
Katedra Metod Matematycznych i Statystycznych
Uniwersytet Przyrodniczy w Poznaniu
sanna6@wp.pl

62. **Zbigniew Szkutnik**
Wydział Matematyki Stosowanej
Akademia Górniczo-Hutnicza w Krakowie
szkutnik@agh.edu.pl

63. **Magdalena Szymkowiak**
Wydział Automatyki, Robotyki i Elektrotechniki
Politechnika Poznańska
magdalena.szymkowiak@put.poznan.pl

64. **Paweł Teisseyre**
Instytut Podstaw Informatyki
Polska Akademia Nauk, Warszawa
Wydział Matematyki i Nauk Informacyjnych
Politechnika Warszawska
teisseyrep@ipipan.waw.pl

65. **Grzegorz Wyłupek**
Instytut Matematyki
Uniwersytet Wrocławski
wylupek@math.uni.wroc.pl

66. **Aleksander Zaigrajew**
Wydział Matematyki i Informatyki
Uniwersytet Mikołaja Kopernika w Toruniu
alzaig@mat.umk.pl

Część VI

Indeks

Indeks